PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47

A2

(11) International Publication Number: **WO 99/06548**

(43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/IB98/01222

(22) International Filing Date: 31 July 1998 (31.07.98)

(30) Priority Data:
08/905,135 1 August 1997 (01.08.97) US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; 24, rue Royale, F–75008 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DUMAS MILNE EDWARDS, Jean–Baptiste [FR/FR]; 8, rue Grégoire–de–Tours, F–75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F–94100 Saint–Maur (FR). LACROIX, Bruno [FR/FR]; 93, route de Vourles, F–69230 Saint–Genis Laval (FR).

(74) Agents: MARTIN, Jean–Jacques et al.; Cabinet Régimbeau, 26, Avenue Kléber, F–75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: 5'ESTs FOR NON TISSUE SPECIFIC SECRETED PROTEINS

(57) Abstract

The sequences of 5'ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5'ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5'ESTs. The 5'ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5'ESTs. The 5'ESTs may also be used to design expression vectors and secretion vectors.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts containing the 5′ EST sequence are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, *Biotechniques*, **13**: 124-131, 1992). Therafter, the resulting phagemids containing the 5′ EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5′ EST sequence. Alternatively, protocoles such as the Gene Trapper kit (Gibco BRL) may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5′ EST sequence are identified by colony PCR or colony hybridization.

Using any of the above described methods in section III, a plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the mature protein remaining after the signal peptide is cleaved off may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5′ ESTs

Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions thereof, such as cDNAs encoding the mature protein, may be used to express the encoded secreted proteins or portions thereof as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

EXAMPLE 30

Expression of the Proteins Encoded by the Genes Corresponding to 5'ESTS or Portions Thereof

To express the proteins encoded by the genes corresponding to 5′ ESTs (or portions thereof), full length cDNAs containing the entire protein coding region or extended cDNAs containing sequences adjacent to the 5′ ESTs (or portions thereof) are obtained as described

in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767, incorporated herein by this reference.

The cDNA cloned into the expression vector may encode the entire protein (*i.e.* the signal peptide and the mature protein), the mature protein (*i.e.* the protein created by cleaving the signal peptide off), only the signal peptide or any other portion thereof.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5′ ESTs or the nucleic acids described above. First, the methionine initiation codon for the gene and the polyA signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a polyA signal, this sequence can be added to the construct by, for example, splicing out the polyA signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained

by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned with the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1 containing a poly A signal and prepared for this ligation (blunt/BglII).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 μg/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein encoded by the extended cDNA

Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate 5' EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5' EST, extended cDNA, or portion thereof.

Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5' EST or a portion thereof are compared

to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be compared to the proteins expressed in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies, the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be β-globin or a nickel binding polypeptide. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the β-globin

gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (*Basic Methods in Molecular Biology*, Davis, Dibner, and Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* Express™ Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5′ ESTs, extended cDNAs, or fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be appreciated that a plurality of proteins expressed from these cDNAs may be included in a panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

EXAMPLE 31

Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5′ ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 466 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 17..127
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.4
seq LWRLLLWAGTAFQ/VX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```
AACTCAGGAC AACGCT ATG GCT GAG CCT GGG CAC AGC CAC CAT CTC TCC GCC   52
                  Met Ala Glu Pro Gly His Ser His His Leu Ser Ala
                      -35                 -30

AGA GTC AGG GGA AGA ACT GAG AGG CGC ATA CCC CGG CTG TGG CGG CTG   100
Arg Val Arg Gly Arg Thr Glu Arg Arg Ile Pro Arg Leu Trp Arg Leu
-25                 -20                 -15                 -10

CTG CTC TGG GCT GGG ACC GCC TTC CAG GTG RMC CAG GGA MSG GRA CCG   148
Leu Leu Trp Ala Gly Thr Ala Phe Gln Val Xaa Gln Gly Xaa Xaa Pro
                -5                  1                   5

GAG CTT CAS GCC TGC AAA GAG TCT GAG TAC CAC TAT GAG TAC ACG GCG   196
Glu Leu Xaa Ala Cys Lys Glu Ser Glu Tyr His Tyr Glu Tyr Thr Ala
         10                  15                  20

TGT GAC AGC ACG GGT TCC AGG TGG AGG GTC GCC GTG CCG CAT ACH YCG   244
Cys Asp Ser Thr Gly Ser Arg Trp Arg Val Ala Val Pro His Thr Xaa
     25                  30                  35

GGC CTG TGC ACC AGC CTG CCT GAC CCC GTC AAG GGC ACC GAG TGC TSN   292
Gly Leu Cys Thr Ser Leu Pro Asp Pro Val Lys Gly Thr Glu Cys Xaa
 40                  45                  50                  55

NTC TCC TGC AAC GCC GGG GAG TTT CTG GAT ATG AAG GAC CAG TCA TGT   340
Xaa Ser Cys Asn Ala Gly Glu Phe Leu Asp Met Lys Asp Gln Ser Cys
                60                  65                  70

NNG CCA TGC GCT GAG GGC CGC TAC TCC CTC GGC ACA GGC ATT CGG TTT   388
Xaa Pro Cys Ala Glu Gly Arg Tyr Ser Leu Gly Thr Gly Ile Arg Phe
            75                  80                  85

GAT GAG TGG GAT GAG CTG CCC CAT GGC TTT GCA GCC TCT CAG CCA ACA   436
Asp Glu Trp Asp Glu Leu Pro His Gly Phe Ala Ala Ser Gln Pro Thr
         90                  95                  100
```

```
TGG AGC TGG ATG ACA GTG CTG CTG AGT CAC                                  466
Trp Ser Trp Met Thr Val Leu Leu Ser His
    105                 110
```

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 4..78
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.1
seq QACLLGLFALILS/GK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```
AAC ATG ACA GCA GAT CCG CGG AAG GGC AGA ATG GGA CTC CAA GCC TGC      48
    Met Thr Ala Asp Pro Arg Lys Gly Arg Met Gly Leu Gln Ala Cys
    -25                 -20                 -15

CTC CTA GGG CTC TTT GCC CTC ATC CTC TCT GGC AAA TGC AGT BAC AGC      96
Leu Leu Gly Leu Phe Ala Leu Ile Leu Ser Gly Lys Cys Ser Xaa Ser
-10                 -5                  1                   5

CCG GAG CCC GAC CAG CGG AGG ACG CTG CCC CCA GGC TGG GTG TCC CTG     144
Pro Glu Pro Asp Gln Arg Arg Thr Leu Pro Pro Gly Trp Val Ser Leu
            10                  15                  20

GGC CGT GCG GAC CCT GAG GAA GAG CTG AGT CTC ACC TTT GCC CTG AGA     192
Gly Arg Ala Asp Pro Glu Glu Glu Leu Ser Leu Thr Phe Ala Leu Arg
        25                  30                  35

CAG CAG AAT GTG GAA AGA CTC TCG GAG CTG GTG CAG GCT GTG TCG GAT     240
Gln Gln Asn Val Glu Arg Leu Ser Glu Leu Val Gln Ala Val Ser Asp
    40                  45                  50

CCC AGC TCT CCT CAA TAC GGA AAA TAC CTG ACC CTA GAG AAT GTG GCT     288
Pro Ser Ser Pro Gln Tyr Gly Lys Tyr Leu Thr Leu Glu Asn Val Ala
55                  60                  65                  70

GAT CTG GTG AGG CCA TCC CCA CTG ACC CCG                             318
Asp Leu Val Arg Pro Ser Pro Leu Thr Pro
                75                  80
```

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: -19..-1
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.4
seq WIFLAAILKGVQC/EV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

```
Met Glu Phe Gly Leu Ser Trp Ile Phe Leu Ala Ala Ile Leu Lys Gly
               -15                 -10                  -5

Val Gln Cys Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Lys
             1               5                  10

Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asp Phe
     15                  20                  25

Thr Asp Ala Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu
 30                  35                  40                  45

Glu Trp Val Ala Asn Ile Xaa Ser Thr Ala Ser Gly Gly Thr Arg Gly
                50                  55                  60

Tyr Ala Ala Pro Val Lys Asp Arg Phe Ile Ile Ser Arg Asp Asp Ser
             65                  70                  75

Arg Asn Thr Leu His Leu Gln Met Asn Gly Leu Lys Xaa Met Thr Gln
         80                  85                  90

Ala Ile Tyr Tyr Cys Ala Thr
     95                  100
```

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 150 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: -37..-1
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.4
seq LWRLLLWAGTAFQ/VX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

```
Met Ala Glu Pro Gly His Ser His His Leu Ser Ala Arg Val Arg Gly
        -35                 -30                 -25
```

```
Arg Thr Glu Arg Arg Ile Pro Arg Leu Trp Arg Leu Leu Leu Trp Ala
-20                 -15                 -10

Gly Thr Ala Phe Gln Val Xaa Gln Gly Xaa Xaa Pro Glu Leu Xaa Ala
-5                  1                   5                   10

Cys Lys Glu Ser Glu Tyr His Tyr Glu Tyr Thr Ala Cys Asp Ser Thr
            15                  20                  25

Gly Ser Arg Trp Arg Val Ala Val Pro His Thr Xaa Gly Leu Cys Thr
        30                  35                  40

Ser Leu Pro Asp Pro Val Lys Gly Thr Glu Cys Xaa Xaa Ser Cys Asn
    45                  50                  55

Ala Gly Glu Phe Leu Asp Met Lys Asp Gln Ser Cys Xaa Pro Cys Ala
60                  65                  70                  75

Glu Gly Arg Tyr Ser Leu Gly Thr Gly Ile Arg Phe Asp Glu Trp Asp
                80                  85                  90

Glu Leu Pro His Gly Phe Ala Ala Ser Gln Pro Thr Trp Ser Trp Met
            95                  100                 105

Thr Val Leu Leu Ser His
        110
```

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 105 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: -25..-1
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.1
seq QACLLGLFALILS/GK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

```
Met Thr Ala Asp Pro Arg Lys Gly Arg Met Gly Leu Gln Ala Cys Leu
-25                 -20                 -15                 -10

Leu Gly Leu Phe Ala Leu Ile Leu Ser Gly Lys Cys Ser Xaa Ser Pro
                -5                  1                   5

Glu Pro Asp Gln Arg Arg Thr Leu Pro Pro Gly Trp Val Ser Leu Gly
            10                  15                  20
```

SEQ ID NO:7

ID AAX41107 standard; cDNA; 466 BP.
XX
AC AAX41107;
XX
DT 17-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:51.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906548-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01222.
XX
PR 01-AUG-1997; 97US-0905135.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153778/13.
DR P-PSDB; AAY12274.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
XX
PS Claim 1; Page 198-199; 824pp; English.
XX
CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12261 to
CC AAY12514, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX
SQ Sequence 466 BP; 87 A; 135 C; 147 G; 84 T; 13 other;

Query Match 28.0%; Score 444; DB 20; Length 466;
Best Local Similarity 97.0%; Pred. No. 1.4e-91;
Matches 450; Conservative 9; Mismatches 4; Indels 1; Gaps 1;

```
Qy     294 ACTCAGGACAACGCTATGGCTGAGCCTGGGCACAGCCACCATCTCTCCGCCAGAGTCAGG 353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 ACTCAGGACAACGCTATGGCTGAGCCTGGGCACAGCCACCATCTCTCCGCCAGAGTCAGG 61

Qy     354 GGAAGAACTGAGAGGCGCATACCCCGGCTGTGGCGGCTGCTGCTCTGGGCTGGGACCGCC 413
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 GGAAGAACTGAGAGGCGCATACCCCGGCTGTGGCGGCTGCTGCTCTGGGCTGGGACCGCC 121

Qy     414 TTCCAGGTGACCCAGGGAACGGGACCGGAGCTTCACGCCTGCAAAGAGTCTGAGTACCAC 473
           |||||||||::|||||::||:||||||||||||:||||||||||||||||||||||||||
Db     122 TTCCAGGTGRMCCAGGGAMSGGRACCGGAGCTTCASGCCTGCAAAGAGTCTGAGTACCAC 181

Qy     474 TATGAGTACACGGCGTGTGACAGCACGGGTTCCAGGTGGAGGGTCGCCGTGCCGCATACC 533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db     182 TATGAGTACACGGCGTGTGACAGCACGGGTTCCAGGTGGAGGGTCGCCGTGCCGCATACH 241

Qy     534 CCGGGCCTGTGCACCAGCCTGCCTGACCCCGTCAAGGGCACCGAGTGCTCCTTCTCCTGC 593
           :||||||||||||||||||||||||||||||||||||||||||||||||:  ||||||||
Db     242 YCGGGCCTGTGCACCAGCCTGCCTGACCCCGTCAAGGGCACCGAGTGCTSNNTCTCCTGC 301

Qy     594 AACGCCGGGGAGTTTCTGGATATGAAGGACCAGTCATGTAAGCCATGCGCTGAGGGCCGC 653
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db     302 AACGCCGGGGAGTTTCTGGATATGAAGGACCAGTCATGTNNGCCATGCGCTGAGGGCCGC 361

Qy     654 TACTCCCTCGGCACAGGCATTCGGTTTGATGAGTGGGATGAGCTGCCCCATGGCTTTGCC 713
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     362 TACTCCCTCGGCACAGGCATTCGGTTTGATGAGTGGGATGAGCTGCCCCATGGCTTTG-C 420

Qy     714 AGCCTCTCAGCCAACATGGAGCTGGATGACAGTGCTGCTGAGTC 757
           |||||||||||||||||||||||||||||||||||||||||||
Db     421 AGCCTCTCAGCCAACATGGAGCTGGATGACAGTGCTGCTGAGTC 464
```

SEQ ID NO:8

ID AAY12274 standard; Protein; 150 AA.
XX
AC AAY12274;
XX
DT 17-JUN-1999 (first entry)
XX
DE Human 5' EST secreted protein SEQ ID NO:305.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9906548-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01222.
XX
PR 01-AUG-1997; 97US-0905135.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153778/13.
DR N-PSDB; AAX41107.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
XX
PS Claim 27; Page 655-656; 824pp; English.
XX
CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12261 to
CC AAY12514, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 150 AA;

Query Match 34.6%; Score 710; DB 20; Length 150;
Best Local Similarity 93.4%; Pred. No. 1.7e-55;
Matches 127; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```
Qy       1 MAEPGHSHHLSARVRGRTERRIPRLWRLLLWAGTAFQVTQGTGPELHACKESEYHYEYTA 60
           |||||||||||||||||||||||||||||||||||||||||||| ||  ||| ||||||||||||
Db       1 MAEPGHSHHLSARVRGRTERRIPRLWRLLLWAGTAFQVXQGXXPELXACKESEYHYEYTA 60

Qy      61 CDSTGSRWRVAVPHTPGLCTSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGT 120
           ||||||||||||||| ||||||||||||||  |||||||||||||| ||||||||||
Db      61 CDSTGSRWRVAVPHTXGLCTSLPDPVKGTECXXSCNAGEFLDMKDQSCXPCAEGRYSLGT 120
```

```
Qy     121 GIRFDEWDELPHGFAS 136
           ||||||||||||||:
Db     121 GIRFDEWDELPHGFAA 136
```

```
ID   AAX41107 standard; cDNA; 466 BP.
XX
AC   AAX41107;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:51.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01222.
XX
PR   01-AUG-1997;   97US-0905135.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153778/13.
DR   P-PSDB; AAY12274.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
XX
PS   Claim 1; Page 198-199; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
```

CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 466 BP; 87 A; 135 C; 147 G; 84 T; 13 other;

Alignment Scores:

Pred. No.:	4.31e-59	Length:	466
Score:	751.00	Matches:	141
Percent Similarity:	94.00%	Conservative:	0
Best Local Similarity:	94.00%	Mismatches:	9
Query Match:	36.63%	Indels:	1
DB:	20	Gaps:	0

US-09-781-880-8 (1-372) x AAX41107 (1-466)

```
Qy       1 MetAlaGluProGlyHisSerHisHisLeuSerAlaArgValArgGlyArgThrGluArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 ATGGCTGAGCCTGGGCACAGCCACCATCTCTCCGCCAGAGTCAGGGGAAGAACTGAGAGG 76

Qy      21 ArgIleProArgLeuTrpArgLeuLeuLeuTrpAlaGlyThrAlaPheGlnValThrGln 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      77 CGCATACCCCGGCTGTGGCGGCTGCTGCTCTGGGCTGGGACCGCCTTCCAGGTGRMCCAG 136

Qy      41 GlyThrGlyProGluLeuHisAlaCysLysGluSerGluTyrHisTyrGluTyrThrAla 60
           |||      |||||||||   ||||||||||||||||||||||||||||||||||||||||
Db     137 GGAMSGGRACCGGAGCTTCASGCCTGCAAAGAGTCTGAGTACCACTATGAGTACACGGCG 196

Qy      61 CysAspSerThrGlySerArgTrpArgValAlaValProHisThrProGlyLeuCysThr 80
           ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db     197 TGTGACAGCACGGGTTCCAGGTGGAGGGTCGCCGTGCCGCATACHYCGGGCCTGTGCACC 256

Qy      81 SerLeuProAspProValLysGlyThrGluCysSerPheSerCysAsnAlaGlyGluPhe 100
           |||||||||||||||||||||||||||||||      ||||||||||||||||||||||
Db     257 AGCCTGCCTGACCCCGTCAAGGGCACCGAGTGCTSNNTCTCCTGCAACGCCGGGGAGTTT 316

Qy     101 LeuAspMetLysAspGlnSerCysLysProCysAlaGluGlyArgTyrSerLeuGlyThr 120
           ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db     317 CTGGATATGAAGGACCAGTCATGTNNGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACA 376

Qy     121 GlyIleArgPheAspGluTrpAspGluLeuProHisGlyPheAlaSerLeuSerAlaAsn 140
           ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db     377 GGCATTCGGTTTGATGAGTGGGATGAGCTGCCCCATGGCTTTGC-AGCCTCTCAGCCAAC 435

Qy     141 MetGluLeuAspAspSerAlaAlaGluSer 150
           |||||||||||||||||||||||||||||
Db     436 ATGGAGCTGGATGACAGTGCTGCTGAGTCA 465
```

ID AAX41107 standard; cDNA; 466 BP.
XX

SEQ ID NO:9

AC AAX41107;
XX
DT 17-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:51.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906548-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01222.
XX
PR 01-AUG-1997; 97US-0905135.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153778/13.
DR P-PSDB; AAY12274.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
XX
PS Claim 1; Page 198-199; 824pp; English.
XX
CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12261 to
CC AAY12514, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX
SQ Sequence 466 BP; 87 A; 135 C; 147 G; 84 T; 13 other;

Query Match 38.3%; Score 429; DB 20; Length 466;
Best Local Similarity 96.9%; Pred. No. 5.5e-116;
Matches 435; Conservative 9; Mismatches 4; Indels 1; Gaps 1;

```
Qy       1 ATGGCTGAGCCTGGGCACAGCCACCATCTCTCCGCCAGAGTCAGGGGAAGAACTGAGAGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 ATGGCTGAGCCTGGGCACAGCCACCATCTCTCCGCCAGAGTCAGGGGAAGAACTGAGAGG 76

Qy      61 CGCATACCCCGGCTGTGGCGGCTGCTGCTCTGGGCTGGGACCGCCTTCCAGGTGACCCAG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::||||
Db      77 CGCATACCCCGGCTGTGGCGGCTGCTGCTCTGGGCTGGGACCGCCTTCCAGGTGRMCCAG 136

Qy     121 GGAACGGGACCGGAGCTTCACGCCTGCAAAGAGTCTGAGTACCACTATGAGTACACGGCG 180
           |||::||:||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     137 GGAMSGGRACCGGAGCTTCASGCCTGCAAAGAGTCTGAGTACCACTATGAGTACACGGCG 196

Qy     181 TGTGACAGCACGGGTTCCAGGTGGAGGGTCGCCGTGCCGCATACCCCGGGCCTGTGCACC 240
           ||||||||||||||||||||||||||||||||||||||||||||||::||||||||||||||
Db     197 TGTGACAGCACGGGTTCCAGGTGGAGGGTCGCCGTGCCGCATACHYCGGGCCTGTGCACC 256

Qy     241 AGCCTGCCTGACCCCGTCAAGGGCACCGAGTGCTCCTTCTCCTGCAACGCCGGGGAGTTT 300
           |||||||||||||||||||||||||||||||||:  ||||||||||||||||||||||
Db     257 AGCCTGCCTGACCCCGTCAAGGGCACCGAGTGCTSNNTCTCCTGCAACGCCGGGGAGTTT 316

Qy     301 CTGGATATGAAGGACCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACA 360
           ||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db     317 CTGGATATGAAGGACCAGTCATGTNNGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACA 376

Qy     361 GGCATTCGGTTTGATGAGTGGGATGAGCTGCCCCATGGCTTTGCCAGCCTCTCAGCCAAC 420
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     377 GGCATTCGGTTTGATGAGTGGGATGAGCTGCCCCATGGCTTTG-CAGCCTCTCAGCCAAC 435

Qy     421 ATGGAGCTGGATGACAGTGCTGCTGAGTC 449
           ||||||||||||||||||||||||||||||
Db     436 ATGGAGCTGGATGACAGTGCTGCTGAGTC 464
```